

Claims

1. A file server system including  
a plurality of file server nodes;  
at least one inter-node connectivity element coupled to said plurality of  
nodes;  
at least one switch coupled to said plurality of nodes and disposed for coupling file server commands to ones thereof;  
said nodes including a set of pairs, each said pair being coupled to a set of  
storage elements and being disposed to control said storage elements in response to said  
file server commands.
2. A system as in claim 1, wherein at least some of said pairs are disposed for failover from a first node to a second node.
3. A system as in claim 1, wherein each said node includes a processor and a memory.
4. A system as in claim 1, wherein  
each said storage element corresponds to one said pair;  
each said storage element is coupled to both nodes in said corresponding  
pair;

1           whereby both nodes in said corresponding pair are equally capable of con-  
2 trolling said storage element.

3  
4           5.     A system as in claim 1, wherein said connectivity element includes a  
5 NUMA network.

6  
7           6.     A system as in claim 1, wherein said file server system is scalable by  
8 addition of a set of pairs of said nodes.

9  
10          7.     A system as in claim 1, wherein said set of storage elements coupled  
11 to at least one said pair includes a RAID storage system.

12  
13          8.     A system as in claim 1, wherein  
14 each pair includes a first node and a second node;  
15 each pair is disposed to receive file server commands directed to either said  
16 first node or to said second node;

17           each pair is disposed when said file server commands are directed to said  
18 first node to execute said file server commands at said first node and to store a copy of  
19 said file server commands at said second node; and

20           each pair is disposed when said file server commands are directed to said  
21 second node to execute said file server commands at said second node and to store a copy  
22 of said file server commands at said first node.

1  
2 9. A system as in claim 8, wherein  
3 each said pair is disposed when said file server commands are directed to  
4 said first node and said first node is inoperable to execute said file server commands at  
5 said second node; and

6 each pair is disposed when said file server commands are directed to said  
7 second node and said second node is inoperable to execute said file server commands at  
8 said first node.

9  
10 10. A system as in claim 1, wherein  
11 each pair is disposed to receive a file server command;  
12 each pair is disposed so that a first node responds to said file server com-  
13 mand while a second node records said file server command; and  
14 each pair is disposed to failover from said first node to said second node.

15  
16 11. A system as in claim 10, wherein  
17 each pair is disposed to receive a second file server command;  
18 each pair is disposed so that said second node responds to said second file  
19 server command while said first node records said file server command; and  
20 each pair is disposed to failover from said first node to said second node.  
21

1           12. A system as in claim 10, wherein said first node controls said storage  
2 elements in response to said file server command while said second node is coupled to  
3 said storage elements and does not control said storage elements in response to said file  
4 server command.

5  
6           13. A method of operating a file server system, said method including  
7 steps for

8           operating a plurality of file server nodes in a set of pairs, each said pair be-  
9 ing responsive to a set of file server commands;

10           coupling said file server commands to said pairs;

11           coupling a set of messages between ones of said nodes in a first said pair  
12 and ones of said nodes in a second said pair.

13  
14           14. A method as in claim 13, including steps for failover from a first  
15 node to a second node, and from said second node to said first node, in each said pair.

16  
17           15. A method as in claim 13, including steps for scaling said file server  
18 by addition of a set of pairs of said nodes.

19  
20           16. A method as in claim 13, including steps for controlling a set of  
21 storage elements corresponding to one said pair from either node in said pair.

1 17. A method as in claim 16, including steps for operating said set of  
2 storage elements according to a RAID storage method.

3  
4 18. A method as in claim 13, including steps for  
5 receiving file server commands directed to either a first node or to a second  
6 node in each said pair;

7 when said file server commands are directed to said first node, responding  
8 to said file server commands at said first node and storing a copy of said file server com-  
9 mands at said second node; and

10 when said file server commands are directed to said second node, respond-  
11 ing to said file server commands at said second node and storing a copy of said file server  
12 commands at said first node.

13  
14 19. A method as in claim 18, including steps for  
15 when said file server commands are directed to said first node and said first  
16 node is inoperable, responding to said file server commands at said second node using  
17 said copy at said second node; and

18 when said file server commands are directed to said second node and said  
19 second node is inoperable, responding to said file server commands at said first node us-  
20 ing said copy at said first node.

21  
22 20. A method as in claim 13, including steps for

1 receiving a file server command at one said pair;  
2 responding to said file server command at a first node while recording said  
3 file server command at a second node; and  
4 failing over from said first node to said second node.

5  
6 21. A method as in claim 20, including steps for  
7 receiving a second file server command at said one pair;  
8 responding to said file server command at said second node while recording  
9 said file server command at said first node; and  
10 failing over from said first node to said second node.

11  
12 22. A method as in claim 20, including steps for controlling said storage  
13 elements in response to said file server command by said first node while said second  
14 node is coupled to said storage elements and does not control said storage elements in re-  
15 sponse to said file server command.

Handwritten signature/initials.